DOCUMENT RESUME

ED 252 565 TM 850 036

AUTHOR McGeever, James M.
TITLE The Decline of Standardized Test Scores in the United States from 1965 to the Present.
INSTITUTION Appalachia Educational Lab., Charleston, W. Va.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
REPORT NO AEL-OP-012
PUB DATE Oct 83
CONTRACT 400-83-0001-P6
NOTE 49p.
PUB TYPE Information Analyses (070) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; Achievement Tests; Aptitude Tests; *Scores; Secondary Education; Standardized Tests; Student Motivation; Test Results; *Trend Analysis
IDENTIFIERS Iowa Tests of Educational Development; Minnesota Scholastic Aptitude Test; Preliminary Scholastic Aptitude Test; Scholastic Aptitude Test; *Test Score Decline

ABSTRACT

Student scores on standardized tests have steadily declined since 1965. Researchers conducted a literature review and completed data analysis to determine the reasons for this decrease, assessing trends for the period from 1965 to 1983. The kinds of tests most commonly used are aptitude, achievement, and tests of personal and social characteristics. Score trends vary from test to test. The reasons for the declining student scores include changes in the composition of test-takers, decreases in the quantity of schooling which students experience, curriculum changes, declines in student motivation, and deterioration of the family system and social environment. These factors, in combination, have contributed to the test score decline for more than the past fifteen years. Efforts to end the decreases must address the curricular and school climate factors identified. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED252565

SCOPE OF INTEREST NOTICE
The Eric Facility has assigned this document for proccesing to: TM
In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should refect their special points of view HE

The Decline of Standardized Test Scores in the United States from 1965 to the Present

James M. McGeever
Educational R & D Specialist

AEL Occasional Paper 012

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy

TM 850 032

Appalachia Educational Laboratory

1031 Quarrier Street / P.O. Box 1348 / Charleston, West Virginia 25325 / 304-347-0400

Occasional Paper Series: AEL's Occasional Paper Series, a product of the NIE-sponsored Regional Services program, reports results of research conducted by Laboratory staff, clients, consultants, or others which may be of interest to educators in the Region. Other papers in the Series are available by contacting the AEL Resource and Referral Center. Documents asterisked (*) are available on a three-week loan basis only.

001: Selected Remediation Programs for Reading and Math: A Guide for State and Local Use

002: The Origin of Ohio Households' Opinions About Public Education

003: Two Tennessee Studies of Kindergarten's Relationship to Grade Retention and Basic Skills Achievement

004: Volume 1--Selected Programs for Reducing Truant and Disruptive Behavior in Schools

Volume 2--Narrative Descriptions of Fourteen Selected Programs for Reducing Truant and Disruptive Behavior in Schools

005: Energy and Education

*006: Survey of State Procedures for the Validation of Educational Programs, Volumes 1-4

007: Improving School Practice: Summary and Proceedings of the 1981 AEL Regional Forum

*008: Community Survey Model for School Districts: Procedural Guide

009: School Effectiveness: Climate, Goals, and Leadership. Summary and Proceedings of a 1982 Regional Exchange Workshop

010: Considering the Research: What Makes An Effective School?

011: Effective Schools are America's Best Bet: Summary and Proceedings of the 1983 AEL Regional Forum

012: The Decline of Standardized Test Scores in the United States from 1965 to the Present

The Decline of Standardized Test Scores in the United States from 1965 to the Present

James M. McGeever
Educational R & D Specialist

October 1983

Educational Services Office
Appalachia Educational Laboratory
Charleston, West Virginia 25325

The Appalachia Educational Laboratory (AEL) is located in Charleston, West Virginia. It's mission is to improve education and educational opportunity for persons who live in the primarily non-urban areas of its member-state Region. AEL accomplishes its mission by:

- documenting educational problems of the Region and sharing the information both with member states and other R & D producers;
- identifying R & D products potentially useful for solving the documented problems and sharing information about these with member states;
- providing R & D technical assistance and training which may include adapting existing R & D products, to lessen documented problems of the Region; and
- continuing to produce R & D projects of national significance in the areas of Lifelong Learning, School/Family Relations, Basic Skills, and others that may be identified.

Information about AEL projects, programs and services is available by contacting the Distribution Center, Appalachia Educational Laboratory, P. O. Box 1348, Charleston, West Virginia 25325.

The project presented or reported herein was performed pursuant to the Regional Services Contract (400-83-0001, P-6) from the National Institute of Education, U.S. Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the Appalachia Educational Laboratory or the National Institute of Education, and no official endorsement by the Appalachia Educational Laboratory or the National Institute of Education should be inferred.

The Appalachia Educational Laboratory, Inc. is an Equal Opportunity/Affirmative Action Employer.

Table of Contents

	Page
Executive Summary	i
Standardized Testing--The Content	1
Major Standardized Tests	3
Overview of the Major Tests	3
Trends by Individual Standardized Tests	8
Reasons for Declining Test Scores	20
Composition of Test-takers	20
Quantity of Schooling	25
Curriculum	28
Student Motivation	35
Social Environment	36

ACKNOWLEDGEMENTS

The author recognizes and extends appreciation to other major contributors to the paper. Dr. Mabel C. Lee, Ms. Marilyn Slack, and Ms. Sevilla Finley, of AEL, contributed substantially to the report preparation.

Executive Summary

Student scores on standardized achievement tests have steadily declined since 1965. The SAT scores have decreased most dramatically during this time. The SAT verbal score dropped from 478 in 1963 to 424 in 1980. The SAT mathematics score declined from 502 in 1963 to 465 in 1980. The decrease was serious with only one-third of the 1980 test-takers doing as well as half of these taking the test in 1963.

Scores on the PSAT, ACT, ITED, and ITBS, also decreased. There are some exceptions. For example, NAEP reading scores improved during the mid-1970's, and the SAT score decline has levelled in recent years.

Researchers have identified a number of possible explanations for the test score decreases. For example, the composition of test takers has changed significiantly. Far more students are staying in school and taking the tests than previously. Also, there were increases in the number of test-takers from groups with traditionally low scores.

Additionally there has been a decline in the quantity of schooling which students experience. The average daily attendance has decreased, while the school retention rate has increased. More students are staying in school, but are attending it less regularly.

The curriculum has also changed significantly. For example, there is less emphasis on the basic skills of reading, writing and mathematics, which are so important for success on the standardized tests.

Some studies indicate that student motivation has decreased during the period of score decline. With less emphasis on the importance of the tests, students are preparing for them less adequately.

A decline in the family system and an increase in divorce has created conditions where families spend less time together talking and reading. Television watching has increased, having a probable negative impact on SAT scores. The overall effect is less time spent at home doing things which would enhance success at school.

All of these factors in combination have contributed to the test score decline for more than the past decade and a half. Whether the score decline will be significantly reversed remains to be seen. In the meantime, efforts to stop the score decreases must address the curricular and school climate factors identified in the text.

THE DECLINE OF STANDARDIZED TEST SCORES IN THE UNITED STATES FROM 1965 TO THE PRESENT

This is the twelfth in AEL's Occasional Paper series which began in 1981. The paper is drawn from a project report which AEL prepared for the Kentucky Department of Education in 1982. AEL conducted a literature review and completed data analysis in assisting the Department's effort to understand declining 10th grade test scores in the State.

This paper is based on the literature review for the Kentucky study. Test score data trends from 1965 to the present on well-known standardized tests are presented. The reasons for the declining test scores are also discussed.

While specific remedies for increasing test scores are not presented here, it is hoped that a careful reading of the report will suggest helpful solutions to the reader. The entire report on which this paper is based may be obtained from AEL.

Standardized Testing--The Context

Educational testing is done for the fundamental purpose of gathering information about students for decision-making. The results of standardized tests can be used to make decisions regarding student placement in a school program or acceptance into an institution of higher education. Individual scores on standardized tests can be compared to the scores of an entire group (norm group) given the same test.

There are three kinds of standardized tests used in schools: (1) aptitude tests, (2) achievement tests, and (3) tests of personal-social characteristics. Each kind of test has a distinct purpose and use in schools.

Some test makers regard the aptitude test as a measure of intelligence, while others point out essential differences between tests of aptitude and intelligence. For example, some say that when a test provides a description of the general mental ability of a population, it is an intelligence test, and when a test measures specific skills, such as mathematics and reading, it is an aptitude test. The California Test of Mental Maturity (CTMM) is an example of an intelligence test which has been used for school-age children. The Scholastic Aptitude Test (SAT) is a test of the aptitude of prospective college students for academic work in postsecondary education.

The purpose of an achievement test is to determine how much a student has accomplished through past education or training. There are achievement tests for nearly every subject in school. The Comprehensive Tests of Basic Skills, for example, measure student achievement in the subject areas of mathematics, reading, language, reference skills, science, and social studies.

Tests of personal-social characteristics, frequently called inventories, provide additional data about the student. These data usually relate to noncognitive student behavior, interests, attitudes, and personality. Examples of this kind of test are the Kuder Occupational Interest Survey and the Minnesota Vocational Interest Inventory.

There are several other kinds of standardized tests. "Speed" tests measure how much work can be completed within a given amount of time. "Power" tests allow students to complete the entire examination, but few people obtain a perfect score because of the difficulty of the items. Standardized tests can be administered to only one person at a time, or, as is usually the case in the nation's schools, they may be given to many people at the same time.

Additionally, standardized tests may be categorized as verbal, non-verbal, and performance tests. The majority of achievement tests in schools are verbal. Non-verbal tests typically use pictures, puzzles, and symbols, rather than words, to provide educational and psychological data on brain-damaged, illiterate, and very young children. Performance tests require students to engage in a physical activity such as designing patterns or playing a musical instrument.

Finally, standardized tests are either criterion-referenced or norm-referenced. In the criterion-referenced test, performance is measured in relation to some specified standard of proficiency (criterion). A norm-referenced test, on the other hand, measures how well a student performs in relation to the scores obtained by a group (the norm group), using the same questions.

In summary, there are a number of kinds of standardized tests used in the nation's schools. While the tests vary in what they measure, they are all used to collect information about students for making decisions regarding selection, prediction, diagnosis and placement of students, for evaluating school effectiveness, and for research (Educational Research Service, Inc., 1976).

Major Standardized Tests

Overview of the Major Tests

Standarized testing has been conducted in American schools since the early part of the century. For example, the College Entrance Examination Board (CEEB) administered college admission tests as early as 1900. Significant revisions in the CEEB's Scholastic Aptitude Test (SAT) took place in 1926 when the familiar multiple-choice SAT was designed, in the

1930's when the test was technically improved, and in 1941 when high school seniors became the standardization group for the test (ERS, 1976). Nearly one million high school seniors take the SAT annually. This represents about one-quarter of the age group and one-half of those going to college.

The two and one-half hour SAT has verbal and mathematics parts with separate scores on a scale from 200 to 800. The mathematics section emphasizes reasoning and measures problem-solving in arithmetic, elementary algebra, and geometry. The verbal section is designed to assess reading skills and understanding of word relationships. This section covers four areas--antonyms, analogies, sentence completion, and reading comprehension. The material for the test is drawn from social, political, scientific, artistic, philosophical, and literary writing (Wirtz, 1977).

Before 1971, high school juniors were given separate administrations of the Preliminary Scholastic Aptitude Tests (PSAT) and the National Merit Scholarship Qualifying Test (NMSQT). In 1971, the College Board and the National Merit Scholarship Corporation designed a single test (PSAT/NMSQT) which would meet the needs of both groups. As does the SAT, this test measures verbal and mathematical comprehension of high school juniors. Administered by Educational Testing Service (ETS) the test has two functions: (1) to help secondary counselors in guiding students in decisions relating to college attendance, and (2) to furnish a Selection Index to facilitate the selection of high school students for National Merit Scholarships. Scores for the verbal and mathematics portions of the PSAT/NMSQT range from 20 to 80, while the Selection Index scale ranges from 60 to 240.

New standardized tests have been introduced into the schools by McGraw-Hill during the past few decades. For example, the California Achievement Test (CAT) was designed in 1940 and has been used throughout the nation since then. A companion test, the Comprehensive Tests of Basic Skills (CTBS) was developed in 1968. The CTBS, which consists of achievement tests in most curricular areas for grades 2-12, is used nationwide and in the Department of Defense overseas dependent schools.

Fifty percent (50%) of all achievement test users employ either the CAT or CTBS. Sixty-six percent (66%) of the 100 largest American school districts use one of the two tests. Six states (Kentucky, New Mexico, South Carolina, Utah, West Virginia, and Wisconsin) conduct statewide testing using CTBS. (Telephone conversation with Vern Dahl, CTB, McGraw-Hill, 1982)

Other major standardized tests in current use include the American College Testing Program (ACT), initiated in 1959. Nearly one million college-bound seniors, mostly in the western, southern, and north central states, take the test annually. ACT measures students' general development in four subject areas--English usage, mathematics usage, social studies, and the natural sciences. Students are required to demonstrate problem-solving and reasoning abilities. ACT emphasizes skills in recognizing the basic elements of correct and effective writing; drawing inferences and conclusions from prose passages; calculating; and interpreting and evaluating data. Scores on the ACT range from a low of 1 to a high of 36. The standard deviation is 5 (ERS, 1976).

The National Assessment of Educational Progress (formerly conducted by the Education Commission of the States, and presently conducted by the Educational Testing Service) and is funded by the National Institute of

Education. NAEP employs basically the same measures which are available to the classroom teacher in periodically assessing the performance of population samples. Testing is done in ten academic areas: art, career and occupational development, citizenship, literature, mathematics, music, reading, science, social studies, and writing. People who take the NAEP test have to complete multiple-choice items as well as demonstrate ability on other "exercises." For example, students may be asked to use science equipment, to write passages, and to work in a group.

Students are sampled nationally from age groups at especially meaningful times in their educational life. Nine-year-olds have just completed their primary education; thirteen-year-olds their elementary education; and seventeen-year-olds their secondary education.

NAEP testing is done periodically with some subjects tested every three years and others every six years. Changes in knowledge, skills, understanding, and attitudes are measured. Test results are reported by geographical regions, age groups, sex, race, size and type of community, and parental education. Scoring procedures are more complicated than the straightfoward multiple-choice standardized tests (ERS, 1976).

The Iowa Tests of Educational Development (ITED) was first used in 1942. It was designed to assess 9th to 12th grade achievement in expression, quantitative thinking, social studies, natural sciences, vocabulary, and use of sources. The test was developed by the Iowa Testing Programs and distributed by Science Research Associates. Standardized nationally, the test has been used by many school systems in the country.

The Iowa Tests of Basic Skills (ITBS) is an achievement test covering vocabulary, reading, mathematics, language, and work study skills for the

elementary grades. The tests were constructed as part of the Iowa Testing Program of the University of Iowa and distributed by Houghton Mifflin. ITBS has been used by many systems in the country.

In addition to the testing done nationwide, some states have conducted their own assessments of student achievement. For example, California has measured the achievement of twelfth-grade pupils through the administration of the Survey of Basic Skills (different from the CTBS and CAT). Test items were provided in the basic skill areas of reading, written expression, and mathematics.

Similarly, the Minnesota Statewide Testing Program administered an aptitude test to all eleventh-grade students in that state. Verbal comprehension and reasoning ability were measured. New Jersey instituted the Educational Assessment Program in 1972 to generate data for educational decision-making at the state and local levels. In the beginning, the fourth and twelfth grades were tested. Later, the fourth, seventh, and tenth grades were administered the test.

The International Educational Achievement Project was a multinational educational research study begun in 1958. The project consisted of three studies, one from 1958 to 1962, one from 1961 to 1966, and another from 1969 to 1978. The United States participated fully in all of the activities from 1958 to the completion of the third study in 1978, providing sufficient numbers of American students to generalize to the student population.

In the United States, testing was done in the subject areas of Science, Reading Comprehension, Literature, French as a Foreign Language and Civic Education. Testing was done for three student populations: (1) students age 10.0 to 10.11 enrolled in full-time schooling, (2) students age 14.0 to

14.11, and (3) students in the terminal year of full-time secondary education programs that were either pre-university programs or programs of the same length.

Trends by Individual Standardized Tests

While there has been a general decline in student scores on standardized tests during the past decade and a half, there has been some variation in this regard. The following section will outline the trends by selected individual tests.

The best documented example of standardized test score decline is that of the Scholastic Aptitude Test. During the 1940's, 1950's, and early 1960's, national average SAT scores improved. But these scores reached a peak in the mid-1960's and then started to decline steadily until this past year.

In 1963, the SAT verbal score stood at 478. That score declined steadily until 1980 when it reached a low of 424. Similarly, the SAT mathematics score was at 502 in 1963. It decreased to 465 in 1980. In 1981, both scores held fast at the 1980 levels.

A study conducted in 1977 by Wirtz and others pointed out that the SAT score decline from 1963 until that year was "serious." The study noted that when the standard deviations are taken into account, the decline meant that only about one-third of the 1977 test takers did as well as half of those taking the SAT in 1963. Additionally, technical analyses of the equating and scaling practices of the tests indicated an "upward drift" of between 8 and 12 points in the scaling of the scores. Further analysis indicated that the declines in the abilities measured by the SAT between 1963 and 1977 were actually 8 to 12 points larger than indicated by the recorded and reported scores.

In 1982, both scores increased for the first time in 19 years. The verbal score increased by 2 points to 426, and the mathematics scores increased 1 point to 466. (See Figure 1).

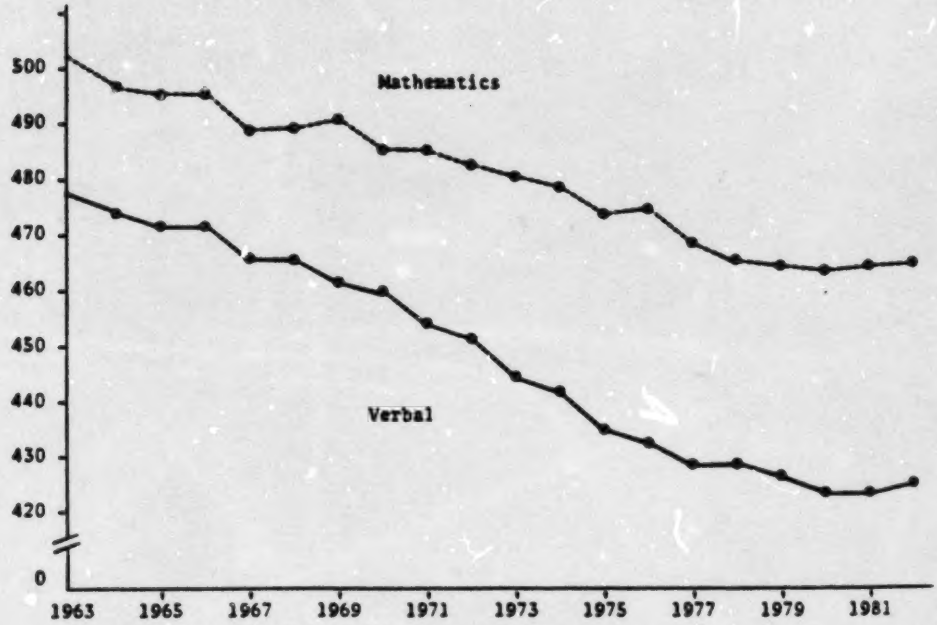

Figure 1

SAT Score Averages, 1963-1982

Source: Chronical of Higher Education, September 1982

Scores on the PSAT, which is taken by high school juniors, did not show a consistent decline as had the SAT scores. Table 1 illustrates the PSAT score data between 1959 and 1975.

Table 1

PSAT Mean Scores for Juniors
1959 to 1975

Year	Verbal	Mathematical
1959	41.2	45.0
1960	40.9	44.8
1961	42.3	46.1
1962	42.9	46.5
1963	43.3	45.6
1964	42.9	44.7
1965	42.2	45.9
1966	42.7	45.0
1967	42.0	44.5
1968	42.6	45.6
1969	42.2	45.0
1970	41.4	46.1
1971*	42.2	45.2
1972*	42.7	46.9
1973*	41.8	45.5
1974*	41.6	45.9
1975*	41.0	45.5

*PSAT/NMSQT
Source: ERS Information Aid: 1976

Results of the American College Testing Program for the decade from 1964-65 to 1974-75 were analyzed. During the ten-year period, the average composite scores declined 1.2 standard score points (Munday, 1976).

The declines varied by subject areas. English scores dropped 1.1 standard score points; mathematics declined 1.5 standard score points; social studies decreased by 2.7 standard score points. Natural science scores, on the other hand, remained constant (Munday, 1976).

The mean composite scores for male and female groups indicated the same downward trend as was exhibited by the combined groups, although men

outperformed women on composite ACT scores (Munday, 1976). Interestingly, the high school grade point averages (GPA's) of seniors who took the ACT showed an upward trend from 1969-70 to 1973-74 with women generally achieving higher high school GPA's then men (Rever and Kojiku, 1975). Finally, while there was a general score decline between 1964-1974 for all regions of the country, the decline was greater in the southern and north central states than in the western states.

The results of the National Assessment in Reading between 1970 and 1980 indicated that nine-year-olds made significant gains, while the performance of thirteen- and seventeen-year-olds remained relatively stable (NAEP, 1980). Thirteen-year-olds gained slightly in literal comprehension while seventeen-year-olds declined slightly in inferential comprehension.

In the nation, nine-year-olds' overall reading performance level rose 3.9%. These students made significant gains in reference skills (4.8%), literal comprehension (3.9%), and inferential comprehension (3.5%). For nine-year-olds, the largest group gains were for black students (9.9%), students who reside in the southeast (7.5%), those who attend schools in rural communities (6.0%) and those who attend schools in disadvantaged-urban communities (5.2%).

Nationally, thirteen-year-olds recorded a significant increase in performance in literal comprehension between 1970 and 1980. Black students registered the only significant overall gain (4.2%) for this group.

For seventeen-year-olds, the performance level declined significantly (2.1%) in inferential comprehension. In most other categories this group's scores held constant.

At each age, three groups--students in the southeast, blacks, and males--continued to perform below the national norm in reading, although they narrowed the gap between them and the nation. Other groups performed at or above the national level.

The National Assessment in Mathematics for the years 1973 and 1978 recorded some interesting trends. Table 2 illustrates the results.

TABLE 2

Changes in Average Performance on
All Mathematics Items Assessed in 1973 and 1978

Ages 9, 13, and 17

	Number of Items	Average Performance 1973	1978	Change in Average Performance
Age 9	55	38%	37%	-1
Age 13	77	53%	51%	-2*
Age 17	102	52%	48%	-4*

*Change is significant at the .05 level.
Source: NAEP, 1979

Students who were 9-years-old demonstrated the lowest decline (38% to 37% in average performance). Still, the decline was close to being significant at the .05 level ($p < .06$). The decline for thirteen-year-olds was somewhat larger (2%) and significant. And seventeen-year-olds' performance diminished appreciably and significantly (NAEP, 1979).

NAEP related the mathematics score declines during this period to other subjects tested:

> It is both interesting and informative to compare the changes in mathematics achievement with the changes observed for other subject areas assessed by National Assessment. In reading, nine-year-olds displayed a slight improvement between 1971 and 1975, while achievement of thirteen- and seventeen-year-olds did not change. Quality of writing, as displayed in written paragraphs, stayed about the same for nine-year-olds between 1970 and 1974, but the quality and coherence of thirteen- and seventeen-year-olds' written expression declined. Basic mechanics of writing--spelling, punctuation and capitalization--showed few changes at any age level. Nine-year-olds' social studies achievement did not change from 1972 to 1976; some declines were evident for thirteen-year-olds and larger drops occurred for seventeen-year-olds. Both teenaged groups also declined in their knowledge and attitudes regarding politics. In science, the achievement level of all three ages decreased between 1970 and 1973; however, between 1973 and 1977, achievement of seventeen-year-olds continued to fall while that of both younger groups stayed about the same. While patterns across subject areas are not completely consistent, some trends are apparent. There is a fairly strong pattern of declines for seventeen-year-olds, the only exception being in the area of reading (1979).

For all three age groups in both 1973 and 1978, students in the northeastern and central regions performed above the national level. Western students were either at or below the nation, and southeastern students were below the national level. In general, all regions declined about the same extent as the nation. At all age levels, males and females recorded identical declines in performance between 1973 and 1978.

Differences in performance by race varied according to age level. For example, nine-year-old blacks showed a significant improvement between 1973 and 1978, while their white classmates recorded a significant decline. Nine-year-olds of Hispanic background demonstrated a constant level of achievement during that period.

For thirteen-year-old whites and Hispanics there was an overall decline in performance between the two testing years. Black thirteen-year-olds did not register a decline.

By age seventeen, all three groups registered significant declines in performance. Whites had declined in average percent of correct response by 3.5%, blacks by 2.6%, and Hispanics by 2.3%.

Scores on the 1973 and 1978 NAEP tests varied according to the type of community involved. For example, disadvantaged students at age nine improved while national performance declined or remained constant. Advantaged-urban students from this age group followed the national pattern of slight decline, as did students from extreme rural communities.

Students age thirteen from disadvantaged-urban communities did not decline in overall average performance, while the nation as a whole did decline. Advantaged-urban and extreme-rural students followed the national pattern of decline.

Seventeen-year-old students from disadvantaged-urban communities evidenced a larger decline than the other two community-type groups. Still, advantaged urban and extreme-rural students who were seventeen-years-old experienced decline.

The achievement patterns for different size or community groups did not change between 1973 and 1978. For all three age groups, students living in big cities were below the national norm while those living in fringes around big cities were above average.

Considering the ITED results for the highly stable student population of Iowa from 1962 through 1974, score declines were evident for grades 9 through 12. The composite scores are illustrated in Figure 2.

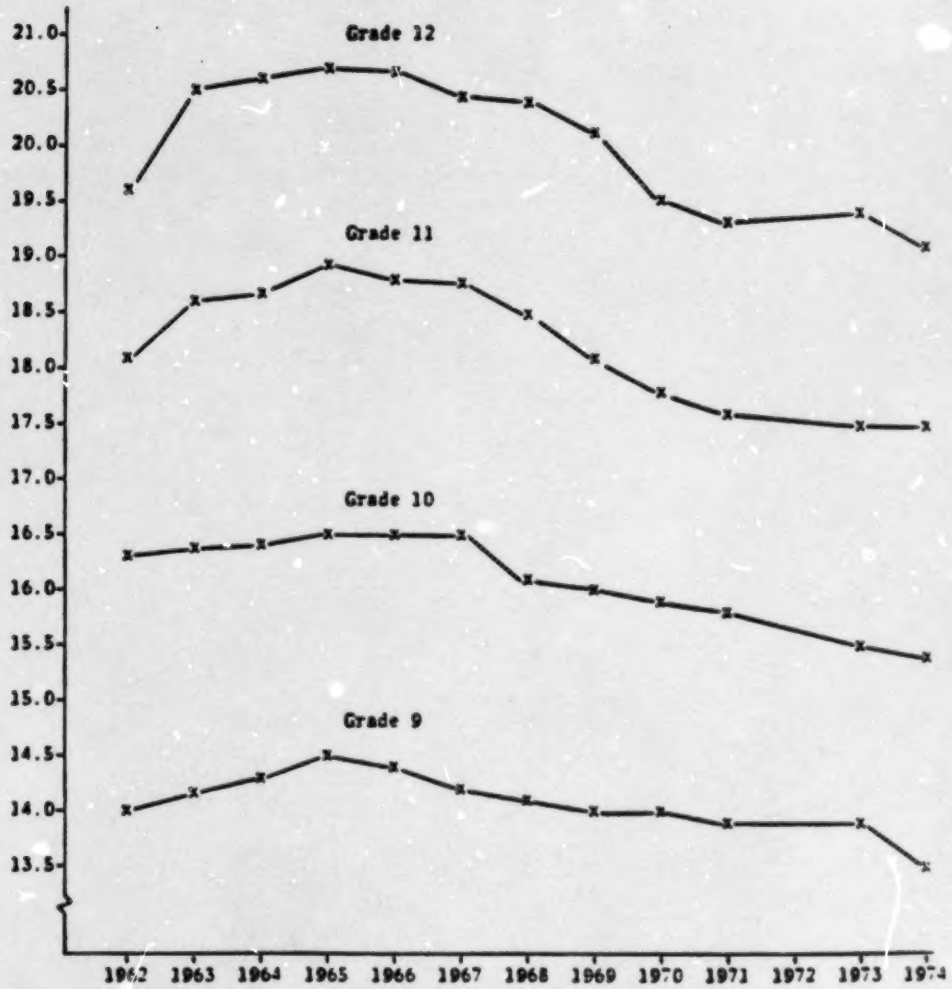

Figure 2

Mean ITED Composite Scores for Grade 9-12, State of Iowa, 1962-74

Source: Harnischfeger and Wiley, 1975

There were score declines evident on all seven sub-tests for all grade levels, grades 9 through 12. However, the declines were most marked for reading materials in literature (Figure 3).

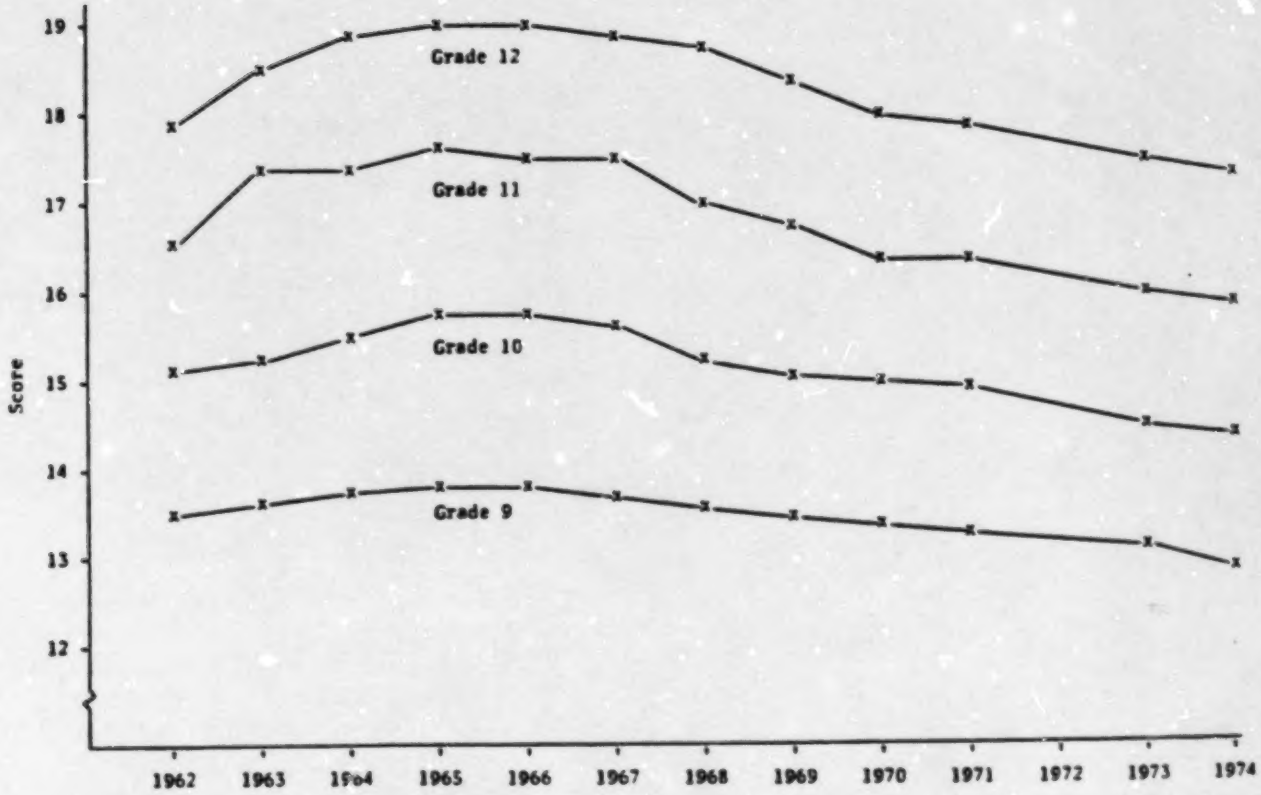

Figure 3

Mean ITED Literature Scores for Grades 9-12, State of Iowa, 1962-1974

Source: Harnischfeger and Wiley, 1975

The general pattern for ITBS scores is one of increase on all sub-scales from 1955 to 1963. However, between 1963 and 1970, national declines were recorded in a majority of sub-scales, including Reading, Language, and Mathematics Skills. Considering ITBS scores for only the State of Iowa between 1965 and 1975, an overall decline was found in all sub-test scores throughout all grade levels.

In four of the ITBS sub-tests (Vocabulary, Reading, Mathematical Problems, and Mathematical Concepts) scores declined between 1965 and 1975. Scores for grades 5 through 8 generally decreased, while scores for grades 3 and 4 did not decrease.

In the California Assessment Program the scores of twelfth graders on the Iowa Tests of Educational Development from 1969 through 1975 were analyzed. Scores showed a continuing decline in all areas. In reading, the scores declined for word meaning, critical comprehension, perceiving cause and effect relationships, and predicting outcomes. In mathematics, there were score declines in fractions, decimals, percents, interpreting graphs, and perceiving and using geometric relationships. These twelfth-grade scores correlated closely with scores earned in previous grades, with socioeconomic status and with parental education.

The Minnesota Statewide Testing Program was administered to all eleventh-grade students in the State. Scores from 1960 through 1973 are shown in Table 3. Between 1968-69 and 1972-73, both the mean and median scores declined. During this period, participation in the Program declined among students of low rank-in-class. Test administrators claimed that the decline would have been more pronounced if the non-participants had contributed to the results. Later analysis indicated that the change was due to a decrease in the frequency of high scores combined with an

increase in the frequency of scores close to the average, not to an increase in the frequency of low scores.

Table 3

MSAT Means, Medians, and Standard Deviations for Minnesota High School Juniors: 1960-1973

Junior Year	MSAT Form	N	Median	Mean	SD
1959-60	A	47,890	NA	30.71	13.91
*					
1961-62	A	45,353	NA	32.38	14.69
*					
1963-64	A	62,434	32.0	34.41	14.73
1964-65	A	61,094	31.8	34.58	14.70
1965-66	A	62,648	32.4	34.71	14.41
1966-67	C	62,783	31.1	32.88	13.61
1967-68	C	67,018	31.1	32.87	13.46
1968-69	C	66,727	31.4	32.96	13.34
1969-70	C	65,830	30.8	32.47	13.19
1970-71	C	66,314	30.3	31.97	12.86
1971-72	C	64,805	30.0	31.53	12.65
1972-73	C	64,628	29.4	31.05	12.39

*Information not available for 1960-61 and 1962-63.
SOURCE: ERS, 1976

The New Jersey Assessment Program recorded important results in tenth-grade scores between 1972 and 1974. The tenth graders' scores were inconsistent. Progress was made in word recognition and literal comprehension skills, but scores declined for higher cognitive skills (cause-effect relationships, drawing conclusions, making inferences). In mathematics, New Jersey students improved in whole number addition and subtraction skills, working with fractions, and in grasping basic algebraic and geometric concepts. On the other hand, the assessment identified weaknesses in concepts of decimals, percentage, measurement, factoring, and

graphing equations. Students from urban center communities scored consistently lower than suburban-urban and suburban-rural in both reading and mathematics.

The third phase of the International Educational Achievement Project measured achievement for a number of countries between 1969 and 1978. In reading comprehension, American ten-year-olds had a higher level of achievement than their counterparts in five countries, but scored lower than their peers in seven other participating countries. Fourteen-year-old American students performed better. The American students out-performed the fourteen-year-old students of ten other countries involved in the international testing, while students from only two countries performed better than students from the United States. By the terminal year of full-time secondary education programs, American students no longer had an achievement advantage over other countries. Students from eleven countries performed better than American students in their final year of secondary school, while students from three tested countries performed below the American achievement level.

A possible explanation for the relatively low achievement of the oldest American students is that a far greater percentage of U.S. students are still enrolled in the final year of secondary school than is the case in other participating countries. The other countries, in short, have a more selected student population during the last year of schooling.

In summary, standardized test scores for the past decade and a half have generally declined. There are some exceptions. For example, NAEP reading scores during the mid-70's have improved, and the decline in SAT scores has levelled during the past three years. SAT scores now show

signs of an upward turn. Still the dominant trend is one of decline in all grades from grade five upward. Score declines are more pronounced in the higher grades and they are more noticeable in verbal than in mathematics achievement.

Reasons for Declining Test Scores

The purpose of this section is to review the factors commonly associated in the literature with the score declines reported in the previous section. The discussion of factors provides some context for understanding the declines. The factors, however, cannot be understood as clear-cut causes for the score declines. In short, it has not yet been established which factors are responsible for the declines and which ones are simply the result (along with the score declines) of other forces producing changes in scores. Most of the factors associated with declining test scores are related to the school itself. A variety of school-level factors will be discussed.

Composition of Test-Takers

Changes in the composition of students who have been taking standardized tests over the past decade and a half seem to be associated with the decline in scores. This seems to apply to both the major achievement and aptitude tests used nationally.

The best documented example of compositional change is in the group of high school students taking the SAT. In 1952, only half of all Americans were staying in school through 12th grade; by 1964, this fraction had increased to two-thirds; by 1970, three-fourths of American students were staying in school. The proportion of these students who were going on to

college was one-fourth in 1952, one-third in 1964, and almost one-half in 1970. More students were staying in school and going on to colleges which required them to take the SAT. Data indicated, as one might expect, that the increasing retention rate was associated with a drop in the average developed ability level of the larger number of students staying in school.

A composite study involving Project Talent, the National Longitudinal Study, and SAT data (based on a comparison of scaled score averages on the equated 1960 and 1972 reading tests of all these groups) indicated that the averages of the SAT-takers dropped by twice as much on these reading tests as did those of high school seniors as a whole. This reflected the increasing percentage of lower scores among those taking the SAT. In 1960, over half (55.4%) of the SAT takers were from the top 20 percent on the Project Talent and National Longitudinal Study reading tests. In 1972, this had dropped to somewhat more than one-third (36.4%) (Wirtz, 1977).

During the early period of the SAT score decline, there were increases in the test-taking population of three groups which traditionally recorded scores far below average: students from lower-socioeconomic-status families, members of minority ethnic groups, and women (for the mathematical portion of the test). These three groups increasingly availed themselves of the SAT during the 1960's and 1970's.

Studies indicate that students from the lowest income families score, on an average, 100 points lower on the SAT (verbal and mathematics) than students from the higher income families. And scores increase consistently from one intermediate income level to the next. 1960 Project

Talent and 1972 National Longitudinal Study data show substantial increases in the percentage of students from lower socioeconomic quarters who went on to colleges (ERS, 1976). Relatedly, there were decreases in the percentages of college bound students among the two higher socioeconomic quarters.

In the early 1960's, very few blacks or other minority ethnic group members took the SAT. Probably no more than 1 percent of SAT-takers were black at that time. By 1972, that percentage increased to about 8 percent. At a time when the Coleman Study (1966) reported that black achievement levels were approximately one standard deviation below those of whites, the SAT results revealed that blacks averaged approximately 100 points below the overall average on the verbal and about 115 points lower on the mathematical part of the SAT (Wirtz, 1977).

A recent report (Chronicle of Higher Education, Oct. 13, 1982) from the College Board indicates that the discrepancies between white and black scores on the SAT have persisted. The 1982 SAT results point to a difference of more than 100 points on both verbal and mathematics portions of the test, with whites outperforming blacks. This year, white students averaged 442 on the verbal section, compared to 332 for blacks. On the mathematics section, whites averaged 482, compared to 362 for blacks. Relatedly, the median family income for whites taking the test was $26,300, more than double the median of $12,500 for blacks. The effects of many years of racial discrimination seem to have figured into the SAT score declines. As noted in the Wirtz Panel Report: "When an increase in the percentage of economically disadvantaged students staying in high school and taking college entrance examinations results in lowering the average scores on these tests, one thing this says is that the national effort to neutralize this kind of disadvantage is still incomplete."

Women and men have always performed equally well on the verbal portion of the SAT, but men have consistently outperformed women on the mathematics part of the test. Between 1960 and 1970 the percentage of women SAT-takers increased from 42.7 percent to 47.5 percent.

Other important compositional changes have occurred in the SAT-taking population. For example, two lower-scoring groups--those who go to two-year colleges and those who do not enter college immediately after high school--increased from 8 percent (1960) to about 15 percent (1972) of the SAT-taking population.

At the same time, two higher-scoring groups--those who go to highly selective liberal arts colleges and those who go to research universities--have decreased as a percentage of SAT-takers. Those SAT students bound for elite liberal arts colleges decreased from 13.2 percent in 1961 to 5.6 percent in 1974. SAT-takers going to research universities diminished from 10.5 percent to 6.4 percent during that period.

Finally, the number of students who take the SAT a second time (averaging 15 to 30 points better in repeating) diminished significantly. This decrease in the number of repeat test-takers had a small effect on the score averages.

The Wirtz Panel concluded that these compositional factors were identifiable with the largest part of the early SAT score declines. The Panel noted that, after fairly careful calculation, these compositional changes are associated with two-thirds to three-fourths of the score decline.

Two additional studies identify other compositional factors in the SAT score declines. Maw (1975) pointed to the increasingly easy access to higher education through financial aid programs, open admissions

policies, and the G.I. Bill as instrumental in encouraging unmotivated and/or unqualified students to take the SAT. McCandless (1975) found that regional shifts in the SAT-taking population were also associated with the score declines. For example, the New England and Mid-Atlantic states usually constituted the highest scoring regional cohort. In 1959, these states accounted for more than half of the SAT takers. By 1968, these high scoring states provided only 40 percent of the SAT-taking population.

Researchers have suggested a number of factors as possibly associated with achievement score declines in the general school population as distinguished from the selective population of SAT-takers. However, data are lacking to make a strong case for most of these explanatory factors.

For example, Harnischfeger and Wiley note that schools, in a response to the baby boom of the fifties, may have sacrificed quality for a time by resorting to double sessions and by hiring disproportionately large numbers of inexperienced teachers. However, they point out that there are insufficient data to make these conclusions.

Harnischfeger and Wiley do identify a potentially promising explanatory factor in increasing school retention rates during the period of test score declines. Data aggregated during the period of score decline are shown in Figure 4.

As illustrated, retention rates rose markedly from 1950 until 1968 when they achieved a consistent level of 750 students staying in school for every 1,000 who entered fifth grade. More and more students staying in schools--students who in previous years would have dropped out--could have had a depressing effect on the test scores.

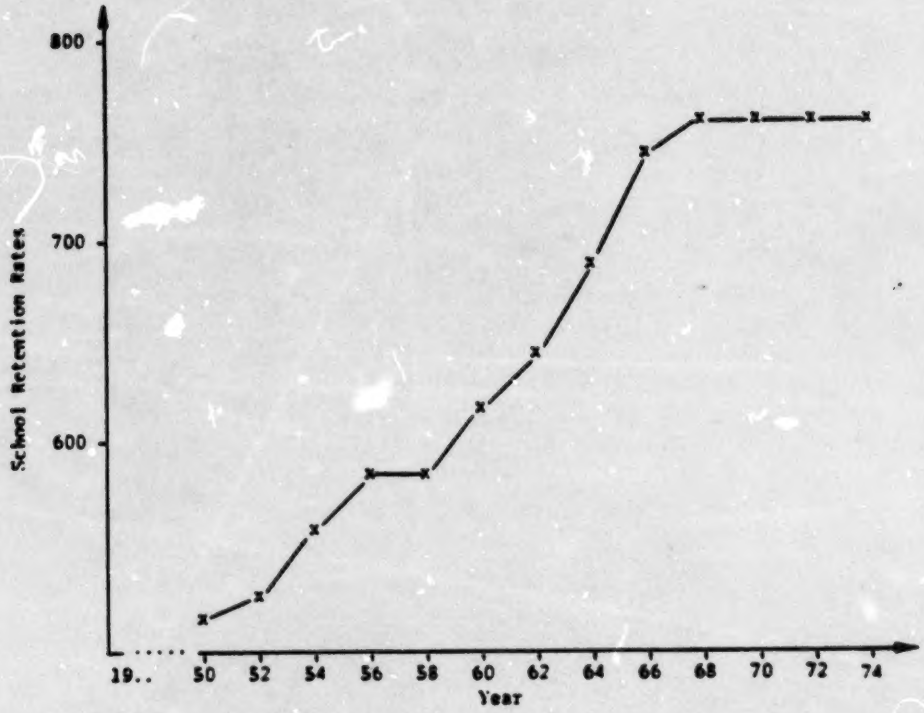

Figure 4

School Retention Rates: High School Graduates per 1,000 Pupils Who Entered 5th Grade

Source: Harnischfeger and Kiley, 1975

Quantity of Schooling

From the early 1950's through the early 1970's the average length of a school year in this country has been rather constant, close to 180 school days per year (See Table 4).

Table 4

Average Daily Attendance and
Average Length of Term
1949/50-1971/72

Year	ADA as % of Enrollment	Average Length of Term of Days	Average Number of Days Attended per pupil Enrolled
1949/50	88.7	177.9	157.9
1951/52	88.7	178.2	156.0
1953/54	88.9	178.6	158.9
1955/56	89.0	178.0	158.5
1957/58	88.7	177.6	157.4
1959/60	90.0	178.[illegible]	160.2
1961/62	90.7	179.1	162.3
1963/64	91.2	179.0	163.2
1965/66	91.4	178.9	163.5
1967/68	91.3	178.8	163.2
1969/70	90.4	178.9	161.7
1971/72	90.2	179.3	161.7

Source: Harnischfeger and Wiley, 1975

However, there was fluctuation in pupil's daily attendance which reached a peak in 1965-66, and then decreased substantially.

Figure 5 illustrates the fluctuations in average daily attendance with steadily increasing school retention rates. These two variables very likely contributed to declining test scores.

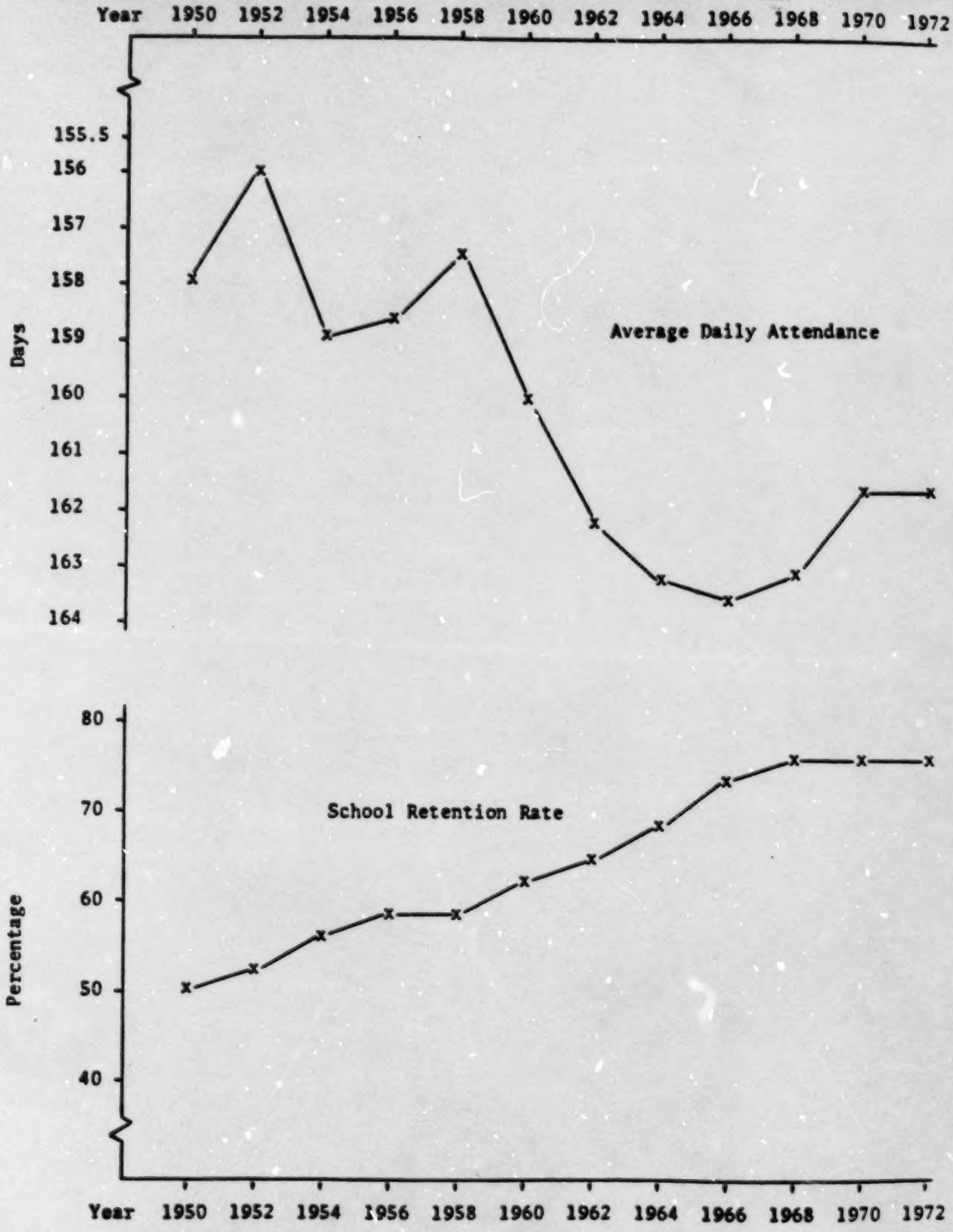

Figure 5

Average Number of Days Attended per Pupil Enrolled and Percentage School Retention Rates over Two-Year Periods from 1958 to 1972

Source: Harnischfeger and Wiley, 1975

Curriculum

In a report prepared for the Wirtz Panel, Wharton (1975) identified a list of hypotheses which have been advanced in letters to the College Board, the Advisory Panel, and in magazine and newspaper articles to explain the test score declines. These hypotheses included:

- neglect of the three Rs (educational innovation at the expense of basic skills);
- increased emphasis on aural-oral communication skills (use of audio visual equipment);
- decreased emphasis on reading and writing skills;
- effects of "new math";
- decline in modern foreign language study (Source: Modern Language Association);
- decreased enrollments in traditional, basic academic courses (Source: Wiley report);
- reading programs that do not emphasize phonetics, vowels, consonants, etc.;
- inferior textbooks (publishers' concern with form rather than content (Sources: newspaper articles; Hudson Institute report);
- stress on independent study;
- open classrooms (where students choose activities in non-graded groups and work with little teacher interference);
- education as entertainment concept (widening variety of nonchallenging, nonacademic courses that entertain more than they educate (Sources: newspaper articles; U.S. News and World Report article); and
- different curriculums followed by different ethnic groups (black students in one California district are enrolled in algebra and geometry courses at half the rate of white students in the same district. Sources: letter from provost, University of California at San Diego).

Few studies provide data to support these hypotheses. However, Harnischfeger and Wiley gathered data which indicated the plausibility of the sixth listed hypotheses--that decreased enrollments in traditional, basic academic courses contributed to the declining test scores. Table 5 illustrates that while course-taking in grades 7 to 12 of American public schools increased every year between 1948 and 1971, students began taking fewer courses during 1972-73.

Table 5

Course Enrollments in Subject Areas and Average Number of Courses per Pupil for Total Pupils Enrolled in Grades 7 to 12 of Public Secondary Schools:

U.S., 1948-49, 1960-61, 1970-71, 1972-73

Subject Area	1948-49		1960-61		1970-71		1972-73	
	Enrollment*	Mean Courses	Enrollment*	Mean Courses	Enrollment*	Mean Courses	Enrollment*	Mean Courses
English-Language Arts	7,099	1.028	12,972	1.106	25,852	1.405	24,079	1.296
Social Sciences	6,982	1.011	11,802	1.001	19,660	1.068	18,899	1.017
Mathematics	4,458	0.645	8,596	0.733	14,137	0.768	13,240	0.713
Natural Sciences	4,031	0.584	7,739	0.660	12,772	0.694	12,475	0.672
Music	2,484	0.360	4,954	0.422	6,559	0.356	6,111	0.329
Foreign Language	1,235	0.179	2,576	0.220	4,729	0.257	4,511	0.243
Physical Education	7,795	1.128	12,082	1.030	22,194	1.206	21,517	1.158
Practical Education	8,731	1.264	14,519	1.237	25,183	1.368	17,743	0.955
Other	111	0.016	106	0.009	233	0.013	9	---
Art	1,220	0.177	2,384	0.203	4,351	0.236	5,116	0.275
Total	44,146	6.392	77,730	6.621	135,670	7.371	123,700	6.658

*Enrollment is in thousands

Source: Patterns of Course Offerings and Enrollments in Public Secondary Schools, 1970-71, and unpublished data. U.S. Department of Health, Education, and Welfare, National Center for Education Statistics.

BEST COPY AVAILABLE

More detailed data are available in Figure 6. These data indicate that, between 1970-71 and 1972-73, the percent of secondary enrollments decreased for each grade level. The drop was substantial, averaging about 14 percent over all grade levels. Enrollment drops were most dramatic at the secondary grade levels. For example, by 1972-73, the percentage of 10th grade enrollments in English courses was down to 80 percent (compared to 95 percent two years before). Eleventh grade English enrollments declined from 88 percent to 65 percent during that two year period.

Figure 7 provides further data on the declining English enrollments by course. It is clear that enrollments decreased in basic and advanced English courses, while they increased in such elective English courses as literature, writing, speech, and drama.

The total proportion of all secondary pupils enrolled in regular English courses decreased from 88.7 percent to 75.2 percent, a drop of 15.2 percent; while the enrollment in other courses only increased 5.2 percent. This total drop of almost 11 percent in English enrollment during a two-year period is a very possible contribution to the verbal score declines of the early 1970's.

In history courses, the overall enrollment percentage remained about the same during 1970-71 to 1972-73. While general history courses decreased in enrollment, the elective history courses increased, taking up much of the slack from the declines.

In the natural sciences, there was a 13 percent drop in general science enrollment in that two-year period. In high school, the decline reached 30 percent. As far as specific sciences were concerned, Biology enrollment suffered a one percent decline, Chemistry a 10 percent

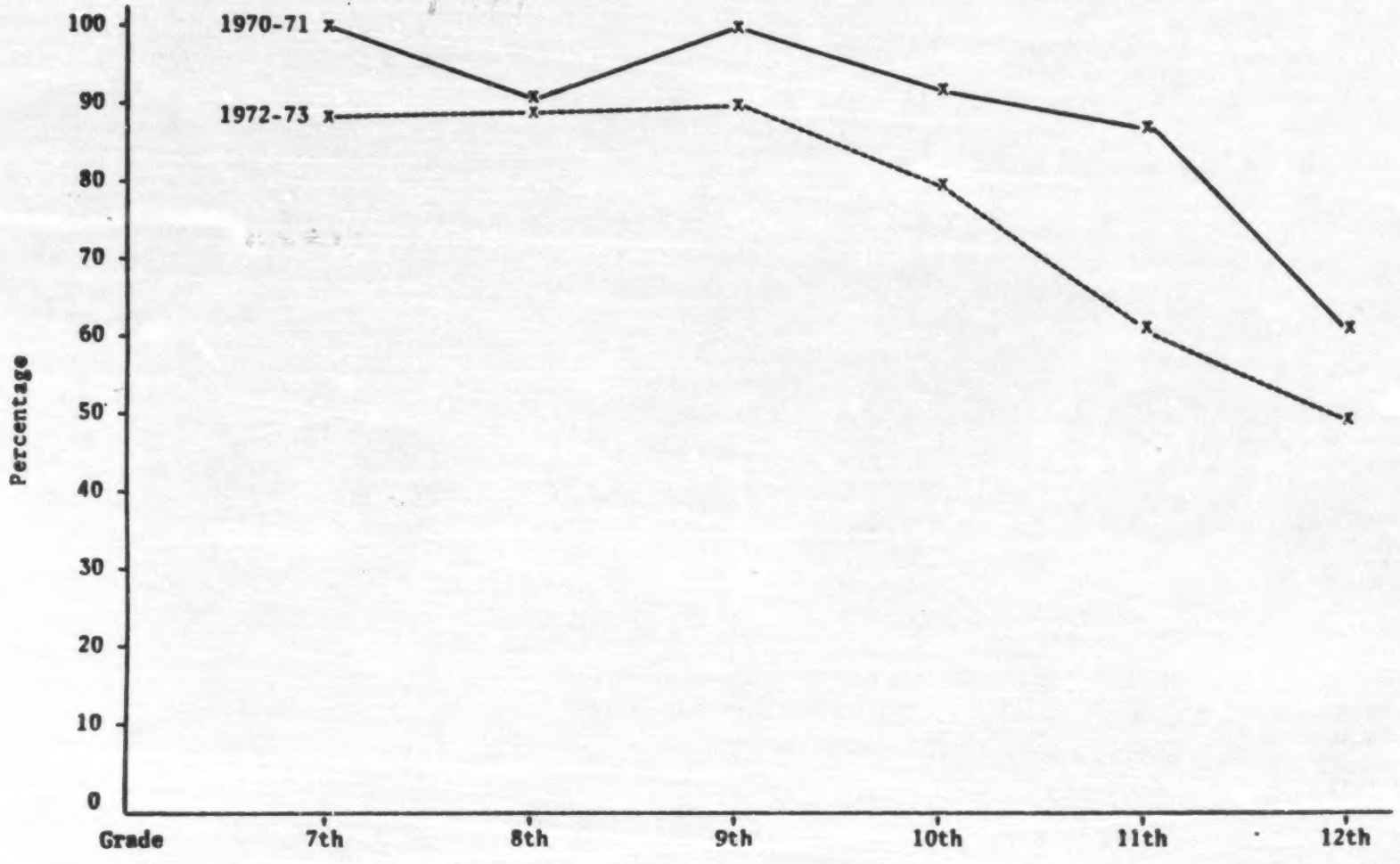

Figure 6

Percent of Secondary Enrollments in English Courses by Grades for the 1970-71 and 1972-73 School Years

Source: Harnischfeger and Wiley, 1975

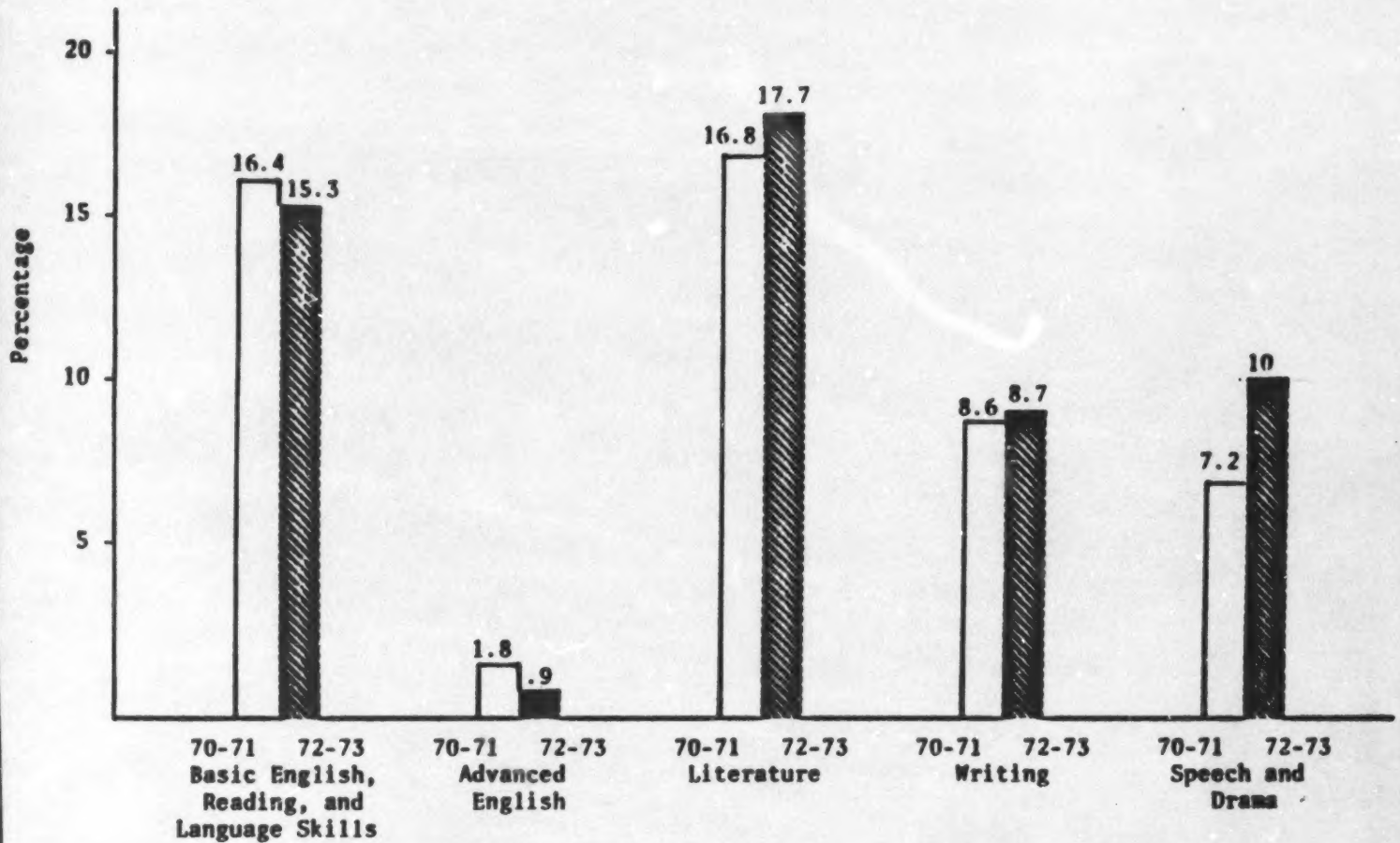

Figure 7

Secondary Enrollments in English Courses by Type of Course for the 1970-71 and 1972-73 School Years

Source: Harnischfeger and Wiley, 1975

decrease, and Physics a 30 percent drop. These declines seemed to be greater for the courses with the more stringent mathematical prerequisties.

For mathematics, an overall 7 percent drop in enrollment took place from 1970-71 to 1972-73. General mathematics enrollments decreased by 15 percent while the advanced mathematics courses increased in enrollment by one-third.

During the early part of the 1970's, then, there was a general enrollment drop in academic courses. The general decline came about because of substantial decreases in general course taking which was not substantially replaced by increases in specialized or elective courses. There was a sizable drop in the proportions of pupils enrolling in the traditional basic courses of the college preparatory curricula: Algebra, Chemistry, and Physics.

Other curricular factors may be associated with the test score declines. Some of the learning approaches used in elementary schools may not adequately prepare students for standardized achievement testing. For example, Babcock reports that the new math emphasizes abstract reasoning at the expense of fundamental skill mastery. He also notes that standardized reading texts with controlled vocabularies are the only literary experiences of children in lower grades and that this inhibits individual initiative in reading.

Education Daily reported the research of Farr, Tuinman, and Rowls which pointed out that while many elementary reading teachers emphasize phonics drills, the standardized tests measure mainly reading comprehension. Test scores begin to decline at this point. The authors stress that there needs to be greater emphasis on total reading behaviors and less concentration on the micro-aspects of reading during the earliest school years.

Relatedly, instructional priorities may vary from the contents of standardized tests. Munday asserts that increasing emphasis has been placed on students' free and creative expression rather than on the rigorous work of teaching verbal skills. In mathematics, he notes more attention is given to understanding number theory than on developing computational skills and facility with algebraic operations.

C. H. Edson (1976) points to "modern" educational ideals which stress individual creativity at the expense of teaching formal grammar, structure, and punctuation as instrumental in the test score declines. Edson suggests a second development in the curriculum which may relate to test score declines. He writes:

> Another popular educational 'philosophy' advocates that all schooling should relate directly to the student's future occupation. Although this type of vocational training, often called Career Education, is not designed to be narrowly technical, critics observe that such training often occurs at the expense of education in the traditionally important areas of reading, writing, and arithmetic.

In a paper presented to the House of Representatives Subcommittee on Elementary, Secondary and Vocational Education, Andrew Porter said that most federally-funded programs which provide help in fundamental skills have focused on the early grades. Much less financial assistance, for reading, he noted, has been given to the secondary schools. As already noted, the most significant score declines have been at the secondary level.

Finally, some educators have maintained that standardized tests are no longer relevant. They say that substantial strides have been made in humanitarian studies, a new respect for students' rights has been cultivated, and greater emphasis has been placed on "human becoming"

through open styles of teaching. These educators suggest that standardized tests do not measure these curricular advances (Wilhelms, 1975).

Student Motivation

Another school-related factor which might serve to explain the decline in standardized test scores is that of student motivation. There is some evidence that a decrease in the level of student motivation is associated with the score decline.

Some studies have discussed this factor in relation to the SAT score declines. For example, Weinman (1977), in a Massachusetts study found that many school administrators thought that current SAT-takers were less motivated than students in the past. A cross-section of high school students confirmed many of the things said by the administrators. The students said that the average high school student did not take the SAT seriously because it did not seem to have much effect on college admissions. (The students noted an exception: their peers who sought admission to the most prestigious colleges did worry about the SAT and took it seriously.)

Students pointed out that their high school and college counselors had told them that universities were de-emphasizing the SAT. Weinman interviewed one student who had six college interviews in which SAT scores were not discussed at all. Another student observed: "I'll get into college anyway, so why worry about the SAT's?"

Interviewees indicated that many students did not prepare adequately for the exam. For example, many of their friends spent the night before

the SAT at parties or at other recreational activities. Students also said that many students did not exert their best effort when responding to the questions.

The apparent low motivation of students to prepare for the SAT was consistent with their claim that colleges no longer emphasized SAT scores. This claim was supported by Weinman's investigation of college admissions procedures. For example, Weinman found that SAT scores were not reviewed by colleges as carefully as they had been previously, that SAT's were taken into account only if they worked in the applicant's favor, and that high school grades (not SAT's) were the most important criteria in evaluating students. Weinman concluded that SAT's were "not as crucial a factor in college admissions as they were 10 years ago."

Social Environment

Some significant changes in the social environment of the country preceded the decline in SAT scores. For example, the decade of the 60's was a period of masked change in the family. The divorce rate began to increase as did the incidence of mothers working outside the home by necessity or choice.

Between 1960 and 1976 the percent of all children under age 18 in this country living with both parents dropped from 89 percent to 80 percent. And the number of children living with either one parent or none increased by 300,000 each year. By 1976, more than half of all women with children of school age (and 40 percent of those with children under 6) were holding either full-time or part-time outside jobs. In short, the opportunities for parents to be "teachers at home" have been steadily decreasing.

Since homes were increasingly becoming places where families seldom had meals together, where parents spent less time reading to, talking to or listening to their children, the positive reactions to reading and language development declined, making the task of teaching reading harder than ever (Babcock, 1974).

As children's time with parents has decreased, their time watching television has proved to be very substantial. By age 16 most children have spent between 10,000 and 15,000 hours watching television, more time than they have spent in school. By first grade, students average between 20 and 35 hours a week in television-viewing. Students' time in television-watching increased by one hour a day between 1960 and 1970. By 1976 television, in the words of the Wirtz Panel, had become "surrogate parent, substitute teacher."

The Panel commented on the significant amount of time spent watching television and its probable impact on SAT scores.

> ...spending 10,000 to 15,000 hours on television's fare means a significant reduction in time, some of which would otherwise have gone into the development of skills and aptitudes measures on college entrance examinations. An unquestionably considerable amount of time at the set used to go into homework and into reading and writing. To call television a thief of time is in a sense to leg the issue. Yet if developing the capacities that are measured by the SAT is taken as a value, the only question is whether television's larceny is petty or grand.

Another factor in the social environment associated with declining test scores is the high standard of living which has made it possible for young people to participate in more varied forms of entertainment and uninterrupted leisure-time activities. The result is that there is less need for students to turn to books for adventure and excitement.

Finally, the youth movements of the 1960's featured a shift away from science and technology to the humanities and "things of the spirit" (ERS, 1976). Youth of that decade expressed disillusionment with the war in Vietnam and alienation from the old establishment. They pursued student rights, formed youth countercultures, rebelled against materialism, sought for new life-styles and a new morality. Youth repudiated achievement in traditional areas of status and privilege, emphasizing instead personal and psychological development. All of these factors may have combined to divert young people from traditional training in academic subjects, possibly contributing to declining performance on the SAT.